NOVEL MIRNA MOLECULES ISOLATED FROM HUMAN

EMBRYONIC STEM CELL

Field of the Invention

5

The present invention relates to novel miRNA molecules, more specifically miRNA molecules isolated from human embryonic stem cell.

Background of the Invention

10

Embryonic stem cells (ESCs) were first derived from mice and are now available from a variety of mammalian systems, including human. They are characterized by nearly unlimited self-renewal in an undifferentiated state under defined culture conditions while retaining differentiation capacity (Evans *et. al.*,

15 *Nature*, 292:154-156, 1981; Martin, *Proc. Natl. Acad. Sci.* U. S. A. 78: 7634-7638, 1981;Smith, *Cold Spring Harbor Laboratory Press, New York*, 2001). During differentiation *in vitro*, embryonic stem cells are able to develop into various kinds of specialized somatic cell types and recapitulate processes of early embryonic development. Thus, embryonic stem cells hold promise as an unlimited source for

20 various clinical and biotechnological applications (Brustle, *Science*, 285:754-756, 1999; Martin, *Proc. Natl. Acad. Sci.* U. S. A. 78: 7634-7638, 1981; Li *et al.*, *Curr. Biol.*, 8:971-974, 1998; Pera *et al.*, *J. Cell. Sci.*, 113:5-10, 2000).

Currently a few molecular regulators are known to participate in the self-renewal and pluripotency of mouse embryonic stem (mES) cells. A POU family

25 transcription factor Oct4, the classical marker of all pluripotent cells, is specifically

expressed in pre-implantation embryos, epiblast, germ cells and pluripotent stem cell lines including embryonic stem cells, embryonic germ (EG) cells and embryonic carcinoma (EC) cells (Palmieri *et al.*, *Dev. Biol.*, 166:259-267, 1994; Yeom *et al.*, *Development*, 122:881-894, 1996). Oct4 plays a critical role in the establishment

5 and maintenance of pluripotent cells in a pluripotent state (Nichols *et al.*, *Cell*, 95:379-391, 1998; Niwa *et al.*, *Nat. Genet.*, 24:372-376, 2000; Pesce *et al.*, *BioEssays*, 20:722-732, 1998). Leukemia inhibitory factor (LIF) can maintain self-renewal of mouse embryonic stem cells through activation of Stat3. Oct4 and Stat3 each interact with various cofactors and regulate the expression of multiple target genes

10 (Niwa *et al.*, *Gene Dev.*, 12:2048-2060, 1998). Two other transcription factors, Sox2 and FoxD3, have been shown to be essential for pluripotency in mice embryos (Avilion *et al.*, *Gene Dev.*, 17:126-140, 2003; Hanna *et al.*, *Gene Dev.*, 16:2650-2661, 2002). More recently, it was found that the homeoprotein Nanog is capable of maintaining self-renewal of mouse embryonic stem cell, independently of LIF/Stat3

15 (Chambers *et al.*, *Cell*, 113:643-655, 2003; Mitsui *et al.*, *Cell*, 113:631-642, 2003).

The first human embryonic stem cell line was established only recently (Thomson *et al.*, *Science*, 282:1145-1147, 1998) and 12 lines are publicly available worldwide (NIH Human Embryonic Stem Cell Registry). Despite their great potential, human embryonic stem cells have not been a prolific source of information.

20 This is mainly due to the technical difficulties in cell culture. Maintaining and expanding human embryonic stem cells require laborious and skill-intensive procedures. Moreover, the population-doubling time of human embryonic stem cells is almost three times longer than that of mouse embryonic stem cells (Amit *et al.*, *Dev. Biol.*, 227:271-278, 2000). There exist apparent differences in the characteristics of

25 human embryonic stem cells compared to mouse embryonic stem cells in many

aspects, including the regulation of self-renewal. Of the regulators found in mice, only a few including Oct4 play similar regulatory roles in human embryonic stem cells. Others such as LIF do not affect human embryonic stem cells in maintaining their self-renewal (Reubinoff *et al.*, *Nat. Biotechnol.*, 18:399-404, 2000). Dissecting

5   the regulatory mechanism in human embryonic stem cells will greatly enhance the understanding of stem cells as well as their application.

Recent advances in small RNA research have implicated microRNAs (hereinafter, referred to as 'miRNAs') as important regulators of development and differentiation. miRNAs constitute a large family of non-coding small RNAs of ~22

10   nucleotides (nt) in length. Our understanding of miRNA function originates from studies of the developmentally regulated miRNAs lin-4 (Olsen and Ambros, *Dev. Biol.*, 216:671-680, 1999; Lee *et al.*, *Cell*, 75: 843-854, 1993; and Wightman *et al.*, *Cell*, 75:855-862, 1993) and *let-7* (Reinhart *et al.*, *Nature*, 403:901-906, 2000) in *Caenorhabditis elegans*. By binding and inhibiting the translation of the target

15   mRNA, the lin-4 and let-7 RNAs play an important role in regulating the timing of larval development. Another example is bantam RNA from *Drosophila melanogaster*, which is expressed in a temporal and tissue-specific manner during development, suppressing apoptosis and stimulating cell proliferation by inhibiting translation of hid mRNA (Brennecke *et al.*, *Cell*, 113:25-26, 2003). Several mouse

20   miRNAs including miR-181 were shown to modulate hematopoiesis (Chen *et al.*, *Science*, 303:83-86, 2003). In plants, miRNAs show a high degree of complementarity to transcription factors that are significant in development (Aukerman and Sakai, *Plant Cell*, 2003; Chen, *Science*, 2003; Llave *et al.*, *Science*, 297:2053-2056, 2002b; Palatnik *et al.*, *Nature*, 425:257-263, 2003 and Rhoades *et al.*, *Cell*, 110:513-520,

25   2002). These miRNAs induce target mRNA cleavage or translational repression,

thereby facilitating plant development and organogenesis.

The expression of miRNAs is often regulated in tissue-specific and developmental stage-specific manners (Aravin *et al., Dev. Cell*, 5:337-350, 2003; Krichevsky *et al., RNA*, 9:1274-1281, 2003; Lagos-Quintana *et al., Science*, 294:853-858, 2002; Pasquinelli *et al., Nature*, 408-86-89, 2000 and Sempere *et al., Dev. Biol.* 259:9-18, 2003), although the regulatory mechanism is still largely unknown. The present inventors have previously shown that miRNAs are transcribed as long primary transcripts (termed pri-miRNAs) (Lee *et al., EMBO J.*, 21:4663-4670, 2002). These primary transcripts are first trimmed into approximately 70 nt stem-loop forms (called pre-miRNAs) by the RNase III type protein, Drosha, in the nucleus (Lee *et al., Nature*, 425:415-419, 2003). Following this initial processing, pre-miRNAs get exported to the cytoplasm by Exportin-5 (Lund *et al., Science*, 303:95-98, 2003 and Yi *et al., Genes Dev.*, 2003) and are subject to a second processing to generate the final product of approximately 22 nt mature miRNAs, by another RNase III type protein Dicer (Grishok *et al., Cell*, 106:23-24, 2001; Hutvagner *et al., Science*, 293:834-838, 2001; Ketting *et al., Genes Dev.*, 15:2654-2659, 2001; and Knight and Bass, *Science*, 293:2269-2271, 2001). This stepwise processing and compartmentalization may allow for the fine regulation of miRNA biogenesis at multiple steps.

More than 300 miRNAs have been reported in diverse eukaryotic organisms so far (Aravin *et al., Dev. Cell*, 5:337-350, 2003; Dostie *et al., RNA*, 9:180-186, 2003; Grad *et al., Mol. Cell.*, 11:1253-1263, 2003; Lagos-Quintana *et al., Science*, 294:853-858, 2001; Lagos-Quintana *et al., Curr. Biol.* 12:735-739, 2002; Lagos-Quintana *et al., RNA*, 9:175-179, 2003; Lai *et al., Genome Biol.*, 4, R42, 2003; Lau *et al., Science*, 294:858-862, 2001; Lee and Ambros, *Science*, 294:862-864, 2001; Lee *et al., Cell*, 75:843-854, 1993; Lim *et al., Genes Dev.*, 2, 2, 2003b; Llave *et al., Plant Cell*,

14:1605-1619, 2002a; Mourelatos *et al.*, *Genes Dev.*, 16:720-728, 2002; Park *et al.*, *Curr. Biol.*, 12:1484-1495, 2002; Reinhart *et al.*, *Nature*, 403-901-906, 2000 and Reinhart *et al.*, *Genes Dev.*, 16:1616-1626, 2002). The majority of miRNA genes were discovered through cDNA cloning from size-fractionated RNA samples.

5   Recently, additional miRNA genes have been identified using computational procedures from the vertebrates, *C. elegans* and *Drosophila*. A bioinformatic study suggested that there exist 200 - 255 miRNAs in humans, accounting for almost 1% of the predicted genes (Lim *et al.*, *Science*, 299, 1540, 2003a). If the prediction is correct, about 100 miRNA genes remain to be identified in humans because 152

10 miRNAs have been reported, of which 109 miRNAs have been experimentally validated (Brennecke and Cohen, *Genome Biol.*, 4, 228, 2003). miRNAs that are expressed only in specific developmental stages or conditions would be difficult to be cloned or validated. However, miRNAs have not been isolated yet from human embryonic stem cells.

15

Detailed Description of the Invention

Therefore, the object of the present invention is to provide novel miRNAs

20 isolated from human embryonic stem cells and uses thereof.

To achieve the object of the present invention, the present invention provides an isolated nucleic acid molecule, comprising

(a) a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-

25   17;

(b) a nucleotide sequence which is the complement of (a);

(c) a nucleotide sequence which has an identity of at least 80% to a sequence of

(a) or (b); or

(d) a nucleotide sequence which hybridizes under stringent conditions to a

5   sequence of (a), (b) or (c),

wherein the nucleic acid molecule was isolated from human embryonic stem cells,

and uses thereof.

The present invention is the first to isolate miRNAs from human embryonic

10   stem cells.

The miRNAs of the present invention may have a nucleotide sequence

selected from the group consisting of SEQ ID NOs: 1-17, as shown in the following

Table 1.

15   [Table 1]

Novel miRNAs isolated from human embryonic stem cells

ID	Sequence	SEQ ID NO
miR-302b*	ACUUUAACAUGGAAGUGCUUUCU	1
miR-302b	UAAGUGCUUCCAUGUUUUAGUAG	2
miR-302c*	UUUAACAUGGGGGUACCUGCUG	3
miR-302c	UAAGUGCUUCCAUGUUUCAGUGG	4
miR-302a*	UAAACGUGGAUGUACUUGCUUU	5
miR-302d	UAAGUGCUUCCAUGUUUGAGUGU	6
miR-367	AAUUGCACUUUAGCAAUGGUGA	7
miR-200c	UAAUACUGCCGGGUAAUGAUGGA	8
miR-368	ACAUAGAGGAAAUUCCACGUUU	9
miR-154*	AAUCAUACACGGUUGACCUAUU	10
miR-369	AAUAAUACAUGGUUGAUCUUU	11
miR-370	GCCUGCUGGGGUGGAACCUGG	12
miR-371	GUGCCGCCAUCUUUUGAGUGU	13
miR-372	AAAGUGCUGCGACAUUUGAGCGU	14

miR-373*	ACUCAAAAUGGGGGCGCUUUCC	15
miR-373	GAAGUGCUUCGAUUUUGGGGUGU	16
miR-374	UUAUAAUACAACCUGAUAAGUG	17

The present invention encompasses nucleic acid molecules which are complementary to the nucleotide sequence of miRNA listed in the above table 1. Moreover, a nucleotide sequence which has an identity of at least 80%, preferably of

5   at least 90% and more preferably of at least 95%, to the nucleotide sequence selected from the group of consisting of SEQ ID NOs: 1-17 or the complementary sequence thereof, is included in the present invention. The term "identity" refers to the degree of sequence identity between two nucleic acid sequences, more particularly to the degree that two bases on the same position precisely corresponds to each other in two

10   aligned sequences. The identity can be determined using a identity search program known in the pertinent art such as BLAST (http://www.ncbi.nlm.nih.gov/BLAST/), FASTA (http://bioweb.pasteur.fr/ seqanal/interfaces/fasta.html) or Smith Waterman Algorithm etc.

Furthermore, the present invention encompasses nucleotide sequences which

15   complementarily bind, that is, hybridize, under stringent conditions, with one selected from the group consisting of the nucleotide sequence of SEQ ID NOs: 1-17; the complementary sequence thereof; and the nucleotide sequence having an identity at least 80% to the sequences. The complementary binding conditions may be general hybridization conditions known in the art. Preferably, it comprises reacting (washing)

20   for 1h in 1×SSC and 0.1% SDS at 45℃ or 48℃, more preferably, for 1h in 0.2×SSC and 0.1% SDS at 50℃. Non-hybridized bases are removed during the reacting.

The isolated nucleic acid molecules of the present invention preferably have a

length of from 18 to 100 nt (nucleotides), and more preferably from 18 to 100 nt. The mature miRNAs out of the inventive nucleic acid molecules usually have a length of 19-24 nt, particularly 21, 22 or 23 nt.   The nucleic acid molecules of the present invention may be also provided as a miRNA precursor molecule (pre-miRNA) which generally has a length of 50-100 nt, preferably 65-95 nt.   It should be noted that the miRNA precursor molecule may be produced by processing of a primary transcript which may have a length of > 100 nt.   The miRNA precursor molecule may have a nucleotide sequence selected from the group consisting of SEQ ID NOs: 84-99, as shown in the following table 2.   It preferably has secondary structure as shown in Fig. 2.

[Table 2]

Nucleotide sequences of miRNA precursors of the present invention

miRNAs	Sequence of miRNA precursors (5' → 3')	Size (nt)	SEQ ID NO
miR-302b* and miR-302b	GUUGGGUGGGCUCCCUUCAACUUUAACAUGGAAGUGC UUUCUGUGACUUUAAAAGUAAGUGCUUCCAUGUUUUA GUAGGAGUGAAUCCAAU	91	84
miR-302c* and miR-302c	GGGAUCCCCUUUGCUUUAACAUGGGGGUACCUGCUG UGUGAAACAAAAGUAAGUGCUUCCAUGUUUCAGUGGA GGUGUCUC	81	85
miR-302a* and miR-302a	CCACCACUUAAACGUGGAUGUACUUGCUUU GAAACUAAAGAAGUAAGUGCUUCCAUGUUUUGGUGAU GG	69	86
miR-302d	AGGGGCCCCCUCUACUUUAACAUGGAGGCACUUGCUG UGACAUGACAAAAAUAAGUGCUUCCAUGUUUGAGUGU GGUGGUUCCU	84	87
miR-367	UGGCUACAGGCCAUUACUGUUGCUAAUAUGCAACUCU GUUGAAUAUAAAUUGGAAUUGCACUUUAGCAAUGGUG AUGGAUUGUUAAGCCA	90	88
miR-200c	GGCGGGGGCCCUCGUCUUACCCAGCAGUGUUUGGGUG CGGUUGGGAGUCUCUAAUACUGCCGGGUAAUGAUGGA GGCCCCUGUC	84	89
miR-368	UUUGGUAUUUAAAAGGUGGAUAUUCCUUCUAUGUUUA UGUUAUUUAUGGUUAAACAUAGAGGAAAUUCCACGUU	86	90

	UUCAGUAUCAAA		
miR-154	UACUUGAAGAUAGGUAUCCGUGUUGCCUUCGCUUUAU UUGUGACGAAUCAUACACGGUUGACCUAUUUUUCAGU A	75	91
miR-369	UUGAAGGGAGAUGACCGUGUUAUAUUCGCUUUAUUGA CUUCGAAUAAUACAUGGUUGAUCUUUUCUCAG	69	92
miR-370	AGACAGAGAAGCCAGGUCACGUCUCUGCAGUUACACA GCUCACGAGUGCCUGCUGGGGUGGAACCUGGUCUGUC U	75	93
MiR-301	CUGCUAACGAAUGCUCUGACUUUAUUGCACUACUGUA CUUUACAGCUAGCAGUGCAAUAGUAUUGUCAAAGCAU CUGAAAGCAG	84	94
miR-371	AGCCUGUGGCACUCAAACUGUGGGGGCACUUUCUGCU CUCUGGUGAAAGUGCCGCCAUCUUUUGAGUGUUACCG CU	76	95
miR-372	UCACCCUGUGGGCCUCAAAUGUGGAGCACUAUUCUGA UGUCCAAGUGGAAAGUGCUGCGACAUUUGAGCGUCAC CGGUGA	80	96
miR-373* and miR-373	ACUGGGAUACUCAAAAUGGGGGCGCUUUCCUUUUUGU CUGUACUGGGAAGUGCUUCGAUUUUGGGGUGUCCCUG U	75	97
miR-296	CCCUUCCAGAGGGCCCCCCCUCAAUCCUGUUGUGCCUA AUUCAGAGGGUUGGGUGGAGGCUCUCCUGAAGGG	72	98
miR-374	UACAUCGGCCAUUAUAAUACAACCUGAUAAGUGUUAU AGCACUUAUCAGAUUGUAUUGUAAUUGUCUGUGUA	72	99

※ The underline represents the nucleotide sequences of miRNAs.

The miRNA precursor molecules may be identified using known methods in the pertinent art, such as, MFOLD program (http://www.bioinfo.rpi.edu/applications/mfold/old/rna) (Zuker et al., Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. Kluwer Academic Publishing, Dordrecht, The Netherlands, 1999).

The nucleic acid molecules of the present invention may be present in single-stranded or double-stranded form. The miRNA is usually a single-stranded molecule, while the pre-miRNA is usually an at least partially self-complementary molecule

capable of forming double-stranded portions (e.g. stem- and loop-structures).

The nucleic acid molecules of the present invention may be selected from RNA, DNA or nucleic acid analog molecules, such as sugar- or backbone-modified ribonucleotides or deoxyribonucleotides. It should be noted, however, that the other
5   nucleic analogs, such as peptide nucleic acids (PNA) or locked nucleic acids (LNA), are also suitable.

The nucleic acid molecules of the present invention may be an RNA- or DNA molecule, which contains at least one modified nucleotide analog, i.e. a naturally occurring ribonucelotide or deoxyribonucleotide is substituted by a non-naturally
10   occurring nucleotide. The modified nucleotide analog may be located for example at the 5'-end and/or the 3'-end of the nucleic acid molecule.

Preferred nucleotide analogs are selected from sugar- or backbone-modified
15   ribonucleotides. It should be noted, however, that also nucleobase-modified ribonucleotides, i.e. ribonucleotides containing a non-naturally occurring nucleobase instead of a naturally occurring nucleobase are suitable: for example, uridines or cytidines modified at the 5-position, such as 5-(2-amino)propyl uridine, 5-bromo uridine; adenosines and guanosines modified at the 8-position, such as 8-bromo
20   guanosine; deaza nucleotides, such as 7-deaza-adenosine; O- and N-alkylated nucleotides, e.g. N6-methyl adenosine. In preferred sugar-modified ribonucleotides the 2'-OH group is replaced by a group selected from H, OR, R, halo, SH, SR, NH_2, NHR, NR_2 or CN, wherein R is C_1-C_6 alkyl, alkenyl or alkynyl and halo is F, Cl, Br or I. In preferred backbone-modified ribonucleotides the phosphoester group
25   connecting to adjacent ribonucleotides is replaced by a modified group, such as

phosphothioate group. It should be noted that the above modifications may be combined.

The nucleic acid molecules of the present invention may be obtained by

5   chemical synthesis methods or by recombinant methods, such as by enzymatic transcription from synthetic DNA-templates or from DNA-plasmids isolated from recombinant organisms. Typically phage RNA-polymerases are used for transcription, such as T7, T3 or SP6 RNA-polymerases.

10   The nucleic acid molecules of the present invention may be inserted to a vector by operably linking to an expression control sequence. The expression means transcription and optionally further processing, resulting in a miRNA-molecule or miRNA precursor molecule as described above. The vector is preferably a DNA-vector such as a viral vector or a plasmid. Particularly the vector may be an

15   expression vector suitable for nucleic acid expression in eukaryotic, more particularly mammalian cells.

The viral vector is not limited to, but includes retrovirus, adenovirus, herpes virus and avipox virus. The retroviral vector is constructed so that non-viral proteins can be produced by the viral vector within the infected cells by the elimination or

20   modification of all the virus genes. The main advantages of the retroviral vector for gene therapy lie in the fact that a quantity of genes are transferred into replicative cells, the genes transferred into cell DNA are accurately integrated and continuous infection does not occur after the gene transfection (Miller, A.D., *Nature*, 1992, 357:455-460). The retroviral vector approved by FDA is constructed using PA317 amphotropic

25   retrovirus package cells (Miller, A.D. and Buttimore, C., *Molec. Cell Biol.*, 6:2895-

2902, 1986). As non-retroviral vectors, there is adenovirus as mentioned above (Rosenfeld, M.A., *et al.*, *Cell*, 68:143-155, 1992; Jaffe, H.A. *et al.*, *Nature Genetics*, 1:372-378, 1992; Lemarchand, *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:6482-6486, 1992). The main advantages of the adenovirus lie in the fact that it can transfer a

5   quantity of DNA fragments (36kb genome) and it is capable of infecting non-replicative cells with a very high titer. Also, the herpes virus can be usefully used for human gene therapy (Wolfe, J.H., *et al.*, *Nature Genetics*, 1:379-384, 1992). Besides, other known suitable viral vectors can be used.

As the plasmid expression vector that can be used in the present invention,

10   there can be used mammal expression plasmids known in the pertinent art. For example, they are not limited to, but typically include pRK5 (European Patent No. 307,247), pSV16B (International Patent Publication 91/08291 A) and pVL1392 (PharMingen).

15   The present invention provides diagnostic or therapeutic applications of the inventive nucleic acid molecules. For example, miRNAs may be detected in biological samples, e.g. in tissue sections, in order to determine and classify certain cell types or tissue types or miRNA-associated pathogenic disorders which are characterized by differential expression of miRNA-molecules or miRNA-molecule

20   patterns. Further, the developmental stage of cells may be classified by determining temporarily expressed miRNA-molecules.

MiRNA-302b*, miR-302b, miR-302c*, miR-302c, miR-302a*, miR-203d and miR-367 of the present invention are also expressed in human embryonic

25   carcinoma stem cells. miR-374 is specifically expressed in human cervix epitheloid

carcinoma cell line (HeLa). Therefore, the nucleic acid molecules are suitable for therapeutic applications. That is, the nucleic acid molecules may be used as modulators or targets of developmental processes or disorders associated with developmental dysfunctions, such as cancer.

5

Moreover, the inventive nucleic acid molecules may be used as a modulator of the expression of genes which are at least partially complementary to a nucleic acid molecule selected from the group consisting of SEQ ID NOs: 1-17. Further, miRNA molecules may act as target for therapeutic screening procedures. That is, inhibition or

10   activation of miRNA molecules might modulate a cellular differentiation process such as apoptosis.

Furthermore, the nucleic acid molecules may be used as starting materials for the manufacture of sequence-modified miRNA molecules, in order to modify the

15   target-specificity thereof, such as an oncogene, a multidrug-resistance gene or another therapeutic target gene. The novel engineered miRNA molecules preferably have an identity of at least 80% to the miRNA (starting material) which has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-17. Further, miRNA molecules can be modified, in order that they are symmetrically processed and then

20   generated as double-stranded siRNAs which are again directed against therapeutically relevant targets.

Furthermore, miRNA molecules of the present invention may be used for tissue reprogramming procedures, e.g. a differentiated cell line might be transformed

25   by expression of miRNA molecules into a different cell type or a stem cell.

For diagnostic or therapeutic applications, the inventive nucleic acid molecules are preferably provided as a pharmaceutical composition. This pharmaceutical composition comprises as an active agent at least one nucleic acid

5   molecule of the present invention and optionally a pharmaceutically acceptable carrier.

The administration of the pharmaceutical composition may be carried out by known methods, wherein a nucleic acid is introduced into a desired target cell *in vitro* or *in vivo*. ·

10

Commonly used gene transfer techniques include transient transfection, microinjection, transduction, cell-fusion, calcium phosphate precipitation, liposome-mediated transfection, DEAE dextran-mediated transfection, polybrene-mediated transfection, electroporation, gene gun and viral methods (Wu *et al.*, *J. Bio. Chem.*,

15   267:963-967, 1992; Wu and Wu, *J. Bio. Chem.*, 263:14621-14624, 1988; Graham *et al.*, *Virol.*, 52:456, 1973; McCutchan *et al.*, *J. Natl. Cancer Inst.*, 41:351, 1968; Chu *et al.*, *Nucl. Acids Res.*, 15:1311, 1987; Fraley *et al.*, *J. Biol. Chem.*, 255:10431, 1980; and Capecchi, *Cell*, 22:479, 1980). A recent addition to this arsenal of techniques for the introduction of DNA into cells is the use of cationic liposomes (Felgner *et al.*,

20   *Proc. Natl. Acad., Sci.*, U. S. A. 84:7413, 1987). Commercially available cationic lipid formulations are e.g. Tfx 50 (Promega) or Lipofectamin 2000 (Life Technologies).

The pharmaceutical composition of the present invention may be in form of a

25   solution, e.g. an injectable solution, a cream, ointment, tablet, suspension or the like.

The composition may be administered in any suitable way, e.g. by injection, by oral,

topical, nasal, rectal application etc.   The carrier may be any suitable pharmaceutical

carrier.   Preferably, a carrier is used, which is capable of increasing the efficacy of

the RNA molecules to enter the target-cells.   Suitable examples of such carriers are

5   liposomes, particularly cationic liposomes.

In addition, the present invention provides a marker specific to human stem

cells, comprising nucleic acid molecules of the present invention.   The human stem

cells may be embryonic stem cells, embryonic carcinoma stem cells or adult neural

10   stem cells.   Preferably, the embryonic stem cells may be undifferentiated embryonic

stem cells.   Specifically, out of the nucleic acid molecules of the present invention,

the nucleic acid molecule having a nucleotide sequence selected from the group

consisting of SEQ ID NOs: 1-10 and SEQ ID NOs: 13-16 is specifically expressed in

undifferentiated human embryonic stem cells.   Therefore, the nucleic acid molecules

15   can be usefully used as a marker specific to undifferentiated human embryonic stem

cells.

The present invention provides also a method for determining the type of stem

cells, comprising the steps of:

20   (a) preparing a RNA sample from the stem cells; and

(b) detecting the expression of a nucleic acid molecule having a nucleotide

sequence selected from the group consisting of SEQ ID NOs: 1-10 and SEQ ID NOs:

13-16, in the prepared RNA sample.

The preparation of total RNA from stem cells was performed according to a

25   method known in the pertinent art.   Preferably, it is performed by using TRIzol

reagent (Gibco BRL).   The expression of nucleic acid molecules of the present

invention in the prepared total RNA sample can be detected by a method known in the

pertinent art.   It is preferably detected by performing northern blot analysis using the

nucleic acid molecules as a probe.   As the probe, the nucleic acid having a nucleotide

5   sequence selected from the group consisting of SEQ ID NOs: 37-41, 43-47 and 51-54

can be used.

        As described above, the determination of stem cell type can be performed by

detecting the expression of one of the nucleic acid molecules using a probe specific to

each nucleic acid molecule.   As for another method, it may also be performed by

10   detecting the expression of a cluster containing miRNAs that are present on same

chromosome.   In this case, it is preferably performed by RT-PCR analysis.   The

primers for RT-PCR analysis may be designed on the basis of nucleotide sequences at

both border of the cluster.   For example, the expression of a cluster containing

nucleic   acid   molecules   that   are   present   on   chromosome   4   (miR-

15   302b*~302b~302c*~302c~302a*~302a~302d~367) can be examined.   In this case, it

may be examined by RT-PCR analysis with a primer set of SEQ ID NOs: 74-75.   The

expression of a cluster containing nucleic acid molecules that are present on

chromosome 19 (miR-371~372~373*~373) can be also examined.   In this case, it

may be examined by RT-PCR analysis with a primer set of SEQ ID NOs: 76-77.

20   The inventive method allows to distinguish between human embryonic stem

cells and mouse embryonic stem cells, and to select human embryonic stem cells or

human embryonic carcinoma stem cells.   Especially, for selecting human embryonic

stem cells, the expression of a nucleic acid molecule having a nucleotide sequence

selected from the group consisting of SEQ ID NOs: 1-7 that is specifically expressed

25   in   human   embryonic   stem   cells,   or   a   cluster   containing   the   same   (miR-

302b*~302b~302c*~302c~302a*~302a~302d~367) can be detected. For selecting human embryonic carcinoma stem cells, it may be performed by detecting the expressions of both (a) a nucleic acid molecule (SEQ ID NOs: 1-7) expressed in human embryonic carcinoma stem cells as well as in human embryonic stem cells, or a cluster containing the same; and (b) a nucleic acid molecule (SEQ ID NOs: 8-10 and 13-16) expressed only in human embryonic stem cells, or a cluster containing the same. That is, when the expression of nucleic acid molecule (a) is detected while the expression of nucleic acid molecule (b) is not detected in RNA sample of stem cells, it is determined as embryonic carcinoma stem cells.

The nucleic acid molecules of the present invention and the methods using the same may be utilized as a key factor to identify differences between proliferation- and differentiation mechanisms in human embryonic stem cells and mouse embryonic stem cells. The present invention may provide a regulator to analyze the embryological mechanisms and developmental patterns in human embryonic stem cells and human embryonic carcinoma stem cells using miRNAs that are differentially expressed in these stem cells.

Moreover, it can be determined whether human embryonic stem cells are differentiated using the nucleic acid molecules of the present invention. Therefore, the present invention provides a method for determining the differentiation of human embryonic stem cells, comprising,

(a) preparing a RNA sample from the stem cells; and

(b) detecting the expression of a nucleic acid molecule having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 1-10 and SEQ ID NOs: 13-16, in the prepared RNA sample.

The preparation of total RNA from the stem cells and detection of the expression of nucleic acid molecules of the present invention in the prepared total RNA may be performed as described above.   The nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 37-41, 43-47 and 51-54

5   can be used as the probe.   The method may be performed by detecting the expression of miRNAs on chromosome 4 (miR-302b*~302b~302c*~302c~302a*~302a~302d~367) or a cluster containing the same, or the expression of miRNAs on chromosome 19 (miR-371, miR-372, miR-373* and miR-373) or a cluster containing the same, by northern blot analysis or RT-PCR.   Most preferably, it may be

10   performed by detecting the expression of miRNAs on chromosome 19 (miR-371, miR-372, miR-373* and miR-373) or the cluster containing the same, by northern blot analysis or RT-PCR.

In the embodiment of the present invention, the present inventors have cloned

15   36 small RNAs with known characteristics of miRNAs from human embryonic stem cells.   16 miRNAs of them were identical to previously reported miRNAs from various mammalian adult tissues and cell lines.   Of the remaining 20 miRNAs, 3 miRNAs are identical to those cloned from mouse embryonic stem cells (Houbaviy *et al., Dev. Cell*, 5:351-358, 2003).   Comparison to the study on mouse embryonic stem

20   cells-specific miRNAs revealed interesting similarities between the two miRNA pools from human and mouse embryonic stem cells.   Many of the embryonic stem cells-specific miRNA genes are highly related to each other and organized as clusters. The miR-302b, miR-302c and miR-302d on chromosome 4 appear to be the close homologues of miR-302 that was cloned from mouse embryonic stem cells (Houbaviy

25   *et al., Dev. Cell*, 5:351-358, 2003).

The present inventor found that there are additional miRNA-like sequences related to miR-302 (see Table 3). Mouse miR-302 and its related sequences also form a gene cluster on mouse chromosome 3. Another set of miRNAs, miR-371, miR-372, miR-373* and miR-373 on chromosome 19, is the human homologues of

5   mouse miR-290, miR-291-s, miR-291-as, miR-292-s, miR-292-as, miR-293, miR-294 and miR-295 expressed in mouse embryonic stem cells (Houbaviy et al., Dev. Cell, 5:351-358, 2003). It is intriguing that two human embryonic stem cells-specific miRNA clusters are conversed in the mouse genome. Although the numbers of the homologous genes are different and the sequences are variable in human and mouse

10   clusters, which may implicate divergence of the conserved regulatory pathways, these conserved miRNAs are likely to play central roles in the regulation of mammalian embryonic stem cells.

It should be noted, however, that a considerable proportion of the cloned miRNAs from human and mouse embryonic stem cells are different from each other.

15   7 out of 20 miRNAs identified in the present invention do not have apparent homologues among those cloned from mouse embryonic stem cells. Conversely, 5 out of 15 novel miRNAs from mouse embryonic stem cells (Houbaviy et al., Dev. Cell, 5:351-358, 2003) do not have related nucleotide sequences of the present invention. These results may implicate fundamental differences between the

20   regulatory networks in human and mouse embryonic stem cells.

In another embodiment of the present invention, the expression patterns of 36 miRNAs cloned from human embryonic stem cells were examined. The miRNAs isolated from human embryonic stem cells can be classified into four groups.

25   (1) miRNAs that are expressed in human embryonic stem cells as well as in

human embryonic carcinoma stem cells; miR-302b*, miR-302b, miR-302c*, miR-302c, miR-302a*, miR-302a, miR-302d and miR-367 (shadowed in bright gray in Table 2 and Fig. 2). These miRNAs may have conserved roles in mammalian pluripotent stem cells. All the other miRNAs, except for miR-302a, are novel miRNAs that are newly provided by the present invention.

(2) miRNAs that are expressed specifically in human embryonic stem cells but not in other cells including human embryonic carcinoma stem cells; miR-200c, miR-368, miR-154*, miR-371, miR-372, miR-373* and miR-373 (shadowed in dark gray in Table 3 and Fig. 2). These miRNAs may have functions specific to human embryonic stem cells. It would be interesting to dissect the molecular basis for the differences between the two pluripotent stem cells; human embryonic stem cells and human embryonic carcinoma stem cells. All these miRNAs are novel miRNAs that are newly provided by the present invention.

(3) miRNAs that are rare in human embryonic stem cells but abundant in HeLa and STO cells; let-7a, miR-301, miR-374, miR-21, miR-29b and miR-29. Only miR-374 out of these miRNAs is novel miRNA which is newly provided by the present invention. These stage-specific miRNAs may play roles in the regulation of development and differentiation, like let-7 in *C. elegans*.

(4) The last class consists of miR-16, miR-17-5p, miR-19b, miR-26a, miR-92, miR-103, miR-130a and miR-222. These are expressed in most tested cell lines so they may contribute to basic cellular functions (see Fig. 2).

Specially, of miRNAs provided in the present invention, miR-200c, miR-368, miR-154*, miR-302b*~302b~302c*~302c~302a*~302a~302d~367 cluster on chromosome 4 and miR-371~372~373*~373 cluster on chromosome 19 are expressed specifically in undifferentiated human embryonic stem cells (see Fig. 2 and Fig. 4).

5 The expression pattern of the miRNA cluster on chromosome 19 is particularly interesting. The expression of this cluster becomes evident more rapidly than that of Oct4, which is the earliest marker for human embryonic stem cells known so far (see Fig. 4), and it is detected only during undifferentiation of embryonic stem cells. Also, when their differentiation is started, it is rapidly decreased. Thus, it is tempting

10 to speculate that these miRNAs may be the primary regulators of embryonic stem cell maintenance or differentiation, which act before other known factors including Oct4. These miRNAs may define the very early stage of embryonic development that has not been recognized before.

15 Moreover, the present inventors found that the expression of miR-371~372~373*~373 cluster (chromosome 19) and miR-302b*~302b~302c*~302c~302a*~302a~302d~367 cluster (chromosome 4) were detected in human adult neural stem cells (see Fig. 6). These miRNAs molecules may be involved in regulatory mechanisms of proliferation and differentiation of human adult neural stem cells.

20

For most miRNAs, RNA from only one side of the miRNA precursor is typically cloned or detected on Northern blot. In the present invention, four miRNA genes (miR-302b, miR-302c, miR-302a and miR-373 hairpins) yielded small RNAs corresponding to both strands of the stems (see Table 3 and Fig. 1). However, the

25 frequencies of clones of the two opposite strands are not equal, indicating that there is

a certain degree of asymmetry in the abundance of these miRNAs. In the case of miR-302b and miR-302c, the ratios between the cloning frequencies of the two sides were 22.5:1 and 10:1, respectively. According to the standard nomenclature (Ambros *et al.*, *RNA*, 9:277-279, 2003), the less abundant miRNA of each pair was

5   designated with an asterisk mark. Recent studies using siRNA duplexes demonstrated that this asymmetry comes from asymmetric degradation of the opposing strand following Dicer processing. According to this, the strand with the less stable 5' end has a better chance to survive (Khvorova *et al.*, *Cell*, 11 5:209-216, 2003; Schwarz *et al.*, Cell, 115:199-208, 2003). Sixty-nine percent of novel miRNA

10   precursors provided by the present invention are clearly in agreement with this, indicating that this rule is generally valid although the mechanism of strand selection for miRNA may be more complex than that for siRNA.

Brief Description of the Figures

15

Fig. 1 shows predicted structures of miRNA precursors. RNA secondary structure prediction was performed using MFOLD (version 3.1) and manually refined to accommodate G/U wobble base pairs in the helical segments. The miRNA sequences are underlined.

20   Fig. 2 is northern blot analysis results of miRNAs cloned from human embryonic stem cells. MiRNAs expressed in embryonic stem cells as well as in embryonic carcinoma stem cells are shadowed in bright gray with black border, while miRNAs expressed specifically in embryonic stem cells are shadowed in black.

Fig. 3 shows multiple sequence alignment of the genomic DNA segments

25   corresponding to the miRNA clusters of the present invention. The positions of

dominantly expressed mature miRNAs are shown in box and the positions of weakly

expressed mature miRNAs are underlined.   Conserved residues are indicated with

asterisks.

Fig. 4 is RT-PCR analysis results of the specific expression of the miRNA

5   clusters of the present invention in undifferentiated human embryonic stem cells.

Fig. 5 is schematically depicted genomic organization of miRNA gene

clusters of the present invention.

Fig. 6 is RT-PCR analysis results of the specific expression of the miRNA

clusters of the present invention in human adult neural stem cells.

10

Examples

The present invention will be further described in detail by the examples.

However, the following examples solely illustrate the invention; the matter of

15   the invention should not be construed to be limited thereto.

<Reference example 1>

Culture of embryonic Stem cells

Human embryonic stem cells, SNU-hES3 (Seoul National University, Korea)

20   and MIZ-hES1 (Mizmedi hospital, Korea) were maintained in DMEM/F12 (Gibco

BRL) supplemented with 20 % (v/v) serum replacements (Gibco BRL), penicillin

(100 IU/ml,Gibco BRL) and streptomycin (100 μg/ml, Gibco BRL), 0.1   mM

nonessential amino acids (NAA, Gibco BRL), 0.1 mM Mercaptoethanol (Sigma) and

4 ng/ml basic FGF (R&D).   Media were changed daily.   Human embryonic stem

25   cell colonies were cultured on a feeder layer of mouse STO (ATCC CRL-1503) cells

pre-treated with mitomycin C (Sigma) and were manually detached and transferred onto new STO feeders every 5-6 days. HS-3 mouse embryonic stem cells (Postech) were grown under standard condition (Evans and Kaufman, *Nature*, 292: 154-156, 1981).

5

<Reference example 2>

Differentiation of human embryonic stem cells

To prepare embryoid bodies (EBs), whole colonies of human embryonic stem cells were detached by glass pipette, transferred onto petri-dishes coated with pluronic

10   F-127 (Sigma), and incubated for 10 days. The media for EB were identical to the media for human embryonic stem cell except that it lacked bFGF. Every two days, media were changed using a pipette. To further differentiate EBs made from SNU–hES3, they were plated onto tissue culture plates coated with poly-L-ornithin (0.01% (v/v))/fibronectin (5 g/ml (w/v)). Cells were further incubated for 5 days in N2

15   supplement medium containing 20 ng/ml bFGF and the medium was changed daily. Confluent cells were manually detached, and then pipetted using yellow tips and transferred onto new plates coated with poly-L-ornithin/fibronectin. Cells were cultured for 5 days in N2 medium containing 20 ng/ml bFGF. When the cells reached confluency, they were trypsinized and split 2:1 or 3:1 at new poly-L-

20   ornithin/fibronectin-cotated plates.

<Example 1>

miRNAs cloning from human embryonic stem cells

<1-1> cDNA library construction from human embryonic stem cells and

25   culture thereof

To identify miRNAs expressed in human embryonic stem cells, two independent cDNA libraries were constructed. Total RNA was prepared from each cell line with TRIzol reagent (Gibco BRL). The cDNA libraries were then constructed by directional cloning method using size fractionated RNA (17-26 nt) from undifferentiated human embryonic stem cells (SNU-hES-3, registered at the Korea Stem Cell Research Center) (Lagos-Quintana *et al. Science*, 294: 853-858, 2001). To validate the undifferentiating status of the human embryonic stem cell, SNU-hES3, the steady-state level of Oct4 mRNA was determined by RT-PCR. The RT-PCR was performed by following method: the first-strand cDNA from the indicated cells was synthesized with SUPERSCRIPT (Gibco BRL) using 2-5 μg of total RNA. The PCR was performed using primers of SEQ ID NO: 37 and 38. To assess the undifferentiating status of the human embryonic stem cells, the expressions of alkaline phosphatase (AP) and the developmental stage-specific embryonic antigens (embryonic cell surface markers), SSEA-1, SSEA-3 and SSEA-4 were checked by immunostaining. As a result, SNU-hES3 cells expressed a high level of Oct4, AP, SSEA-3, and SSEA-4(data not shown). From the result, it was validated that the undifferentiating status of SNU-hES3 cells were maintained.

<1-2> MiRNAs cloning from human embryonic stem cells and bioinformatics analysis

MiRNAs were cloned from the cDNA libraries constructed in the Example <1-1> using the method of Lagos-Quintana et al. (Lagos-Quintana *et al. Science*, 294: 853-858, 2001). Database searches of the cloned miRNAs were performed at the BLAST server (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul *et al., J. Mol. Biol.* 215:403-410, 1990) and ENSEMBL server (http://www.ensembl.org) (Hubbard *et al.,*

Nucleic Acids Res., 30:38-41, 2002).   Sequence alignment between miRNA
sequences were performed by using CLUSTALW (http://www.ebi.ac.uk/clustalw/)
(Higgins and Sharp, *Dev. Cell*, 5:351-358, 1988).

As a result, sequences were obtained for 1,475 small cDNAs that resulted in
733 non-redundant sequences.   Approximately 70% of these sequences corresponded
to the expressed strand of loci that encode longer, previously identified coding or
noncoding RNAs such as tRNAs and rRNAs.

To distinguish miRNAs from degradation products or small interfering RNAs
(siRNAs), the present inventors evaluated the ability of RNA containing the clones to
fold into stem-loop, that is, the secondary structure of the RNA by using the MFOLD
program   (http://www.bioinfo.rpi.edu/applications/mfold/old/rna)   (Zuker *et al.*,
Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A
Practical Guide. Kluwer Academic Publishing, Dordrecht, The Netherlands, 1999).
Thirty-six RNAs were found in the stems of strong hairpin structures (see below Table
1 and Fig. 1).

[Table 3]

The miRNAs cloned from human embryonic stem cells.

ID[a]	Sequence[b]	Observation[c]		Size (nt)	Chromosome	Conservation[d]	Expression[e]	SEQ ID NO.
		1st	2nd					
miR-302b*	ACUUUAACAUGGAAGUGCUUUCU	1	1	23	4	Mm	00011	1
miR-302b	UAAGUGCUUCCAUGUUUUAGUAG	9	36	23	4	Mm	00011, 000221	2
miR-302c*	UUUAACAUGGGGGUACCUGCUG	1		22	4	Mm	00011	3
miR-302c	UAAGUGCUUCCAUGUUUCAGUGG	7	3	23	4	Mm	00012	4
miR-302a*	UAAACGUGGAUGUACUUGCUUU	3	3	22	4	Mm	00012	5
miR-302a	UAAGUGCUUCCAUGUUUUGGUGA	6	17	23	4	Mm, Rn	00023, 001223	18
miR-302d	UAAGUGCUUCCAUGUUUGAGUGU	1	12	23	4	Mm	00023,	6

							000112	
miR-367	AAUUGCACUUUAGCAAUGGUGA	2	3	22	4	Mm, Rn	00011	7
miR-200c	UAAUACUGCCGGGUAAUGAUGGA		4	23	12	Mm	00010, 000110	8
miR-368	ACAUAGAGGAAAUUCCACGUUU	1		22	14	Mm	00010, 000110	9
miR-154*	AAUCAUACACGGUUGACCUAUU		1	22	14	Mm	000110	10
miR-369	AAUAAUACAUGGUUGAUCUUU	1		21	14	Mm	no signal	11
miR-370	GCCUGCUGGGGUGGAACCUGG		1	21	14	Mm	00S000	12
miR-301	CAGUGCAAUAGUAUUGUCAAAGC	1		23	17	Mm, Rn, Fr	21111	19
miR-371	GUGCCGCCAUCUUUUGAGUGU		2	21	19	Mm	002120	13
miR-372	AAAGUGCUGCGACAUUUGAGCGU		1	23	19	Mm	000230	14
miR-373*	ACUCAAAAUGGGGGCGCUUUCC		1	22	19	Mm	00SSS0	15
miR-373	GAAGUGCUUCGAUUUUGGGGUGU		1	23	19	Mm	000120	16
miR-296	AGGGCCCCCCUCAAUCCUGU		1	21	20	Mm	00SSS	20
miR-374	UUAUAAUACAACCUGAUAAGUG	2		22	X	Mm	11000	17
Sum		35	87					
let-7a-1	UGAGGUAGUAGGUUGUAUAGUU		1	22	9, 11, 17, 22		33000	21
miR-16	UAGCAGCACGUAAAUAUUGGCG		2	22	13		22112	22
miR-17-5p	CAAAGUGCUUACAGUGCAGGUAGU		1	24	13		11112	23
miR-19b	UGUGCAAAUCCAUGCAAAACUGA	7	1	23	13, X		111111	24
miR-21	UAGCUUAUCAGACUGAUGUUGAC		4	23	17		32111	25
miR-26a	UUCAAGUAAUCCAGGAUAGGCU		3	22	3		221112	26
miR-29	CUAGCACCAUCUGAAAUCGGUU		1	22	7		12SSSS	27
miR-29b-2	UAGCACCAUUUGAAAUCAGUG		1	21	7		11000	28
miR-92	UAUUGCACUUGUCCCGGCCUG	1	1	21	13, X		11112	29
miR-103	AGCAGCAUUGUACAGGGCUAUG	1	2	22	5		21112	30
miR-124a #	UUAAGGCACGCGGUGAAUGCCA	3	1	22	8		00SSS	31
miR-130a #	CAGUGCAAUGUUAAAAGGGCAU		1	22	11		12223	32
miR-134 #	UGUGACUGGUUGACCAGAGGGG	1		22	14		00SSS0	33
miR-135-2 #	UAUGGCUUUUUAUUCCUAUGUGA		1	23	12		no signal	34
miR-136 #	ACUCCAUUUGUUUUGAUGAUGGA	1		23	14		no signal	35
miR-222	AGCUACAUCUGGCUACUGGGUCUC	1	1	24	X		11SSS1	36
sum		15	21					

[a] : miRNAs that were newly identified in the present invention are listed in the upper panel. The shadow in bright gray indicates the miRNAs expressed in both ES cells (embryonic stem cells) and EC cells (embryonic carcinoma stem cells). The

shadow in dark gray indicates the miRNAs expressed in ES cells but not in other cells including EC cells.

[b] : The longest clone is presented.

[c] : Number of the clones found in each library.

[d] : The presence of homologous stem loops in the mouse (Mm), rat (Rn) and pufferfish (Fr) is indicated. Note that these homologues are only predicted ones based on the genomic sequences.

[e] : Expression patterns determined by Northern blot analysis. Single digit numbers indicate the relative band intensities of given miRNA in different cell lines as shown in Fig. 2 and do not give information about the relative levels of different miRNAs. Five digits represent band intensities from HeLa, STO, mES, SNU-hES1, and hEC, consecutively. Six digits represent band intensities from HeLa, STO, mES, SNU-hES1, Miz-hES3, and hEC, consecutively. S indicates a smear around approximately 22 nt that makes it difficult to judge the expression level of the given miRNA.

Of the 36 cloned RNAs listed in the above Table 3, 16 were identical to previously reported miRNAs from various mammalian adult tissues and cell lines (Table 3, lower panel). Interestingly, the remaining 20 RNAs were represented by the majority (77%) of the clones (122 of 158 clones) (Table 3, upper panel). This presents a striking contrast to a recent study using a similar cloning method, where 91% of approximately 600 clones matched previously identified miRNAs (Lagos-Quintata et al., RNA, 9:175-179, 2003). The primary difference was in the source of RNA. That is, Lagos-Quintata et al extracted RNAs from various tissues of 18.5-week-old adult mice and the human osteoblast sarcoma cell line Saos-2 (Lagos-

Quintata *et al., RNA*, 9:175-179, 2003), whereas the present inventors extracted RNAs from human embryonic stem cells.

From the above results, it could be confirmed that human embryonic stem cells are highly enriched with a distinct set of miRNAs unlike the cells in adult tissues or other cell lines.

It was investigated that 3 miRNAs of the remaining 20 miRNAs corresponded to 3 miRNAs which were recently identified from mouse embryonic stem cells (Houbaviy *et al., Dev. Cell*, 5:351-358, 2003): miR-296 (miR-296), miR-301 (miR-301) and miR-302 (miR-302a). Thus, 17 miRNAs cloned from human embryonic stem cells in the present invention are novel (see the above Table 1).

<Example 2>

The specific expression analysis of novel miRNAs in human embryonic stem cells

To validate the expression of the miRNAs cloned in Example 1, Northern blot analysis was performed using total RNAs from undifferentiated human embryonic stem cells (SNU-hES3, Seoul Nation University), mouse embryonic stem cells (HS-3, Postech) and the human embryonic carcinoma stem cell line (NTERA2, ATCC, CRL-1973, USA).

To confirm the expression of miRNAs from human embryonic stem cells, the present inventors employed another human embryonic stem cell line, MIZ-hES1 (Mizmedi hospital, Korea), that is registered in the NIH Human Embryonic Stem Cell Registry. As negative controls, total RNAs from human cervical carcinoma cell line, HeLa cells and mouse fibroblast cell line, STO feeder cells were used.

Total RNA (100 μg) from each cell line was loaded on a 12.5% denaturing

polyacrylamide gel. The resolved RNA was transferred to a Zeta-Probe GT blotting

membrane (Bio-Rad) for about 12h. Oligodeoxinucleotides labeled at the 5' end

with ^{32}P-ɣ-ATP were used as probes. Prehybridization and hybridization were

carried out using Express Hyb Hybridization Solution (Clontech). The used probes

5   are shown in Table 4 below.

[Table 4]

The probes used in Northern blot analysis

Probe	Sequence (5' → 3')	SEQ ID NO.
miR-302b*	AGAAAGCACTTCCATGTTAAAGT	37
miR-302b	CTACTAAAACATGGAAGCACTTA	38
miR-302c*	CAGCAGGTACCCCCATGTTAAA	39
miR-302c	CCACTGAAACATGGAAGCACTTA	40
miR-302a*	AAAGCAAGTACTACCACGTTTA	41
miR-302a	TCACCAAAACATGGAAGCACTTA	42
miR-302d	ACACTCAAACATGGAAGCACTTA	43
miR-367	TCACCATTGCTAAAGTGCAATT	44
miR-200c	TCCATCATTACCCGGCAGTATTA	45
miR-368	AAACGTGGAATTTCCTCTATGT	46
miR-154*	AATAGGTCAACCGTGTATGATT	47
miR-369	AAAGATCAACCATGTATTATT	48
miR-370	CCAGGTTCCACCCCAGCAGGC	49
miR-301	GCTTTGACAATACTATTGCACTG	50
miR-371	ACACTCAAAAGATGGCGGCAC	51
miR-372	ACGCTCAAATGTCGCAGCACTTT	52
miR-373	ACACCCCAAAATCGAAGCACTTC	53
miR-373*	GGAAAGCGCCCCCATTTTGAGT	54
miR-296	ACAGGATTGAGGGGGGGCCCT	55
miR-374	CACTTATCAGGTTGTATTATAA	56
let-7a-1	ACTATACAACCTACTACCTCA	57
miR-16	CGCCAATATTTACGTGCTGCTA	58
miR-17-5p	ACTACCTGCACTGTAAGCACTTTG	59
miR-19b	TCAGTTTTGCATGGATTTGCACA	60

miR-21	GTCAACATCAGTCTGATAAGCTA	61
miR-26a	AGCCTATCCTGGATTACTTGAA	62
miR-29	AACCGATTTCAGATGGAGCTAG	63
miR-29b-2	CACTGATTTCAAATGGTGCTA	64
miR-92	CAGGCCGGGACAAGTGCAATA	65
miR-103	CATAGCCCTGTACAATGCTGCT	66
miR-124a #	TGGCATTCACCGCGTGCCTTAA	67
miR-130a #	ATGCCCTTTTAACATTGCACTG	68
miR-134 #	CCCCTCTGGTCAACCAGTCACA	69
miR-135-2 #	TCACATAGGAATAAAAAGCCATA	70
miR-136 #	TCCATCATCAAAACAAATGGAGT	71
miR-222	GAGACCCAGTAGCCAGATGTAGCT	72

As shown in Fig. 2, of the 17 newly identified miRNAs in the present invention, 14 miRNAs except for miR-369, miR-370 and miR-374, were specifically expressed in a human embryonic stem cell-specific manner; miR-302b*, miR-302b, miR-302c*, miR-302c, miR-302a*, miR-302d, miR-367, miR-200c, miR-368, miR-154*, miR-371, miR-372, miR-373* and miR-373. Of them, miR-302b*, miR-302b, miR-302c*, miR-302c, miR-302a*, miR-302d and miR-367 were expressed both in human embryonic stem cells and in human embryonic carcinoma stem cells. On the other hands, miR-200c, miR-368, miR-154*, miR-371, miR-372, miR-373* and miR-373 were specifically expressed in human embryonic stem cells. Of the remaining 3 clones, miR-374 was mainly expressed in HeLa and STO cell lines. MiR-369 and miR-370 could not be detected. They might be expressed only at very low levels (data not shown).

It could be confirmed that most of miRNAs provided by the present invention are specifically expressed in human embryonic stem cells. Meanwhile, miR302a that was previously reported in mouse embryonic stem cells as miR-302 (Houbaviy *et al.*,

Dev. Cell, 5:351-358, 2003) was also expressed specifically in mouse embryonic stem (mES) cells, human embryonic stem (hES) cells and human embryonic carcinoma stem (hEC) cells. MiR-296 (miR-296), previously cloned from mouse embryonic stem cells, appeared as smear that made it difficult to judge the specificity of expression. MiR-301 was detectable in the all samples tested.

5

<Example 3>

Gene cluster analysis of miRNA gene

Genomic loci for 12 miRNAs cloned in the Example 1 were found in two gene clusters.

10

Eight miRNA loci (miR-302b, miR-302b*, miR-302c, miR-302c*, miR-302a, miR-302a*, miR-302d and miR-367) are located within an about 700 bp region on chromosome 4. Of the above eight miRNA genes, the remaining miRNAs except for miR-302a are novel miRNAs provided by the present invention. Another four miRNAs (miR-371, miR-372, miR-373 and miR-373*) are found within a 1050 bp region on chromosome 19. Sequence comparison of these miRNAs shows that the miRNAs in a given cluster are highly related (see Fig. 3).

15

Especially, four miRNAs from chromosome 4 cluster (miR-302b, miR-302c, miR302a (miR-302) and miR-302d) are highly homologous to each other. Their sequence similarity is greatest in the 5' portions of the miRNA sequences as is the case with the *lin-4* and *let-7* families. This finding is consistent with the hypothesis that target recognition occurs primarily via 5' sequences (Lai, *Nat. Gent.*, 30:363-364, 2002). These related miRNAs may recognize a consensus target sequence and hence act on the same mRNAs or different mRNAs with conserved binding sites. Therefore, recognition of these miRNA gene families should help in the identification

20

25

of putative mRNA targets. The consensus sequence for these clustered miRNAs is 5'-UAAGUGCUUCCAUGUUUNNGUNN-3' (SEQ ID NO: 73) (see Fig. 3A). While these miRNAs are the most abundant ones in human embryonic stem cells, their murine homologue miR-302 appears to be less abundant in mouse embryonic

5   stem cells (Houbaviy et al., Dev. Cell, 5:351-358, 2003). Interestingly, additional sequences related to this family are found in mouse chromosome 3 (see Table 3). Three putative mouse homologues form stem-loop structures and are in a gene cluster.

Of miRNAs provided by the present invention, miR-371, miR-372 and miR-373 are also found in a cluster on chromosome 19 (see Fig. 3B). Of miRNAs

10   predicted by Houbaviy et al., miR-290, miR-291-s, miR-291-as, miR-292-s, miR-292-as, miR-293, miR-294 and miR-295 have sequence similarity to human miRNAs and are expressed in mouse embryonic stem cells. Mouse miR-291-295 are located as one cluster spanning 2.2 kb (Houbaviy et al., Dev. Cell, 5:351-358, 2003).

It is noted that the miRNAs from the two gene clusters, for instance, miR-

15   302a/miR-302 on chromosome 4 and miR-372 on chromosome 19, are similar to some extent, implicating that they may have originated from a common ancestral miRNA gene.

     <Example 4>

20   **Expression pattern analysis of miRNAs in human embryonic stem cell differentiation**

To examine the expression patterns of the miRNAs during differentiation of human embryonic stem cell, RT-PCR was carried out. Because clustered miRNA genes are generally transcribed into pri-miRNAs (polycistronic primary transcripts)

25   (Lee et al., EMBO J., 21:4663-4671, 2002), the primers were chosen to bind outside

the boundary of the predicted approximately 70 nt stem-loop clusters so that the present inventors could detect pri-miRNAs covering the entire cluster. The primers used in RT-PCR analysis are shown in Table 5 below.

5   [Table 5]

The primers used in RT-PCR analysis

miRNA gene cluster	Primer	Sequence (5' → 3')	SEQ ID NO.
miR-302b*~302b~302c*~302c~302a*~302a~302d~367	Forward	GGGCTCCCTTCAACTTTAAC	74
	Reverse	ATTCTGTCATTGGCTTAACAATCCATCACC	75
miR-371~372~373*~373	Forward	CGATCGCCGCCTTGCCGCAT	76
	Reverse	TGGTTCGTGATGCCCTACTCAAACAGGGAC	77
miR-30a~30a*	Forward	ATTGCTGTTTGAATGAGGCTTCAGTACTTT	78
	Reverse	TTCAGCTTTGTAAAAATGTATC AAGAGAT	79
let-7a-1	Forward	GATTCCTTTTCACCATTCACCCTGGATGTT	80
	Reverse	TTTCTATCAGACCGCCTGGATGCAGACTTT	81
GAPDH	Forward	TGTCATCAATGGAAATCCCATCACC	82
	Reverse	CATGAGTCCTTCCACGATACCAAA G	83

Total RNA was extracted from the human embryonic stem cell lines (SNU-hES3 and Miz-hES1), EBs (embryoid bodies) derived from each cell line (SNU-
10   hES3(EB) and Miz-hES1(EB)), differentiated cells (SNU-hES3 (Dif)) derived from EBs, and Hela cell, respectively. Preparation of EBs from human embryonic stem cell lines was performed as used in Reference Example 2. RT-PCR was then performed according to the same method as used in Example 1 using the primers in Table 5 above.

15   As a result, PCR products of the expected size (707 and 1056 bp, respectively) were detected from two human embryonic stem cell lines (see Fig. 4,

lanes 2 and 5), indicating that these clusters are indeed single transcriptional units. The steady-state levels of primary transcripts from the two gene clusters decreased when human embryonic stem cells developed into EBs (see Fig. 4, lanes 3 and 6). The expression levels were reduced further in differentiated cells derived from EBs (see

5   Fig. 4, lane 4).   These results demonstrate that the clustered miRNAs are expressed specifically in human embryonic stem cells and are rapidly down-regulated during differentiation.

To assess the degree of differentiation, the present inventors determined the

10   steady-state levels of Oct4 mRNAs at differential stages by RT-PCR according to the same method as used in Example 1 (see Fig. 4).   Interestingly, down-regulation of chromosome 19 cluster (miRNA-371-372-373-373*) precedes that of Oct4.   The polycistronic transcripts from the chromosome 4 cluster, but not those from the chromosome 19, were detected in human embryonic carcinoma stem cells (see Fig. 4,

15   lane 7), which is consistent with the results from Northern blot analysis (see Fig. 2).

The control miRNA (miR-30a), which had not been cloned from human embryonic stem cells, was detected in HeLa cells but only barely shown in other cells in accordance with Northern results of Example 2 (see Figs. 2 and 4).   Meanwhile, pri-let-7a-1 transcript was expressed in all of the tested cell lines, although mature let-

20   7a-1 appeared only in HeLa and STO cell lines, as shown in Fig. 2, which suggests that the processing of let-7a-1 may be regulated posttranscriptionally.   It would be of great interest to understand how the expression of miRNAs is regulated during development.

25   <Example 5>

Expression analysis of miRNAs in human adult neural stem cells

<5-1> Culture of adult neural stem cells

The cells obtained from the periventricular zone of 8-12 weeks old embryo telencephalon were stably cultured and suspended at concentration of 5×10^5 cells/ml.

5   The cell suspension was cultured in a tissue-culture dish.   For media, DMEM/F12 as a basal medium and N2 medium supplemented with bFGF (10-20 µg/ml), heparin (8 µg/ml) and EGF (10-20 µg/ml) were used.   Every 5 days, media were changed.

<5-2> RT-PCR analysis

10   Total RNA was prepared from human adult neural stem cells (hNSC) prepared in the above Example <5-1> and the human embryonic stem cell line (Miz-hES1), respectively.   To validate the expression of cluster containing miRNAs of the present invention (miR-302b*~302b~302c*~302c~302a*~302a~302d~367 and miR-371~372~373*~373), RT-PCR analysis was performed using a primer set (SEQ ID

15   NO: 74/75 and 76/77, respectively) as used in Example 1.   The expression of β-actin was validated as a control.

As shown in Fig. 6, miR–302b*~302b~302c*~302c~302a*~302a~302d ~367 cluster on chromosome 4 and miR–371~372~373*~373 cluster on chromosome 19 were expressed specifically in human adult neural stem cell (hNSC) as well as human

20   embryonic stem cell.   It suggests that miRNAs of the present invention play a part in regulatory mechanism of proliferation and differentiation in human adult neural stem cells as well as of human embryonic stem cells.

Industrial Applicability

25

As described above, the novel miRNAs were cloned from human stem cells in the present invention.   The miRNA molecules provided by the present invention can be usefully used as a molecular marker for early developmental stages of undifferentiated human embryonic stem cells.   Also, the miRNA molecules of the present invention may play an important role in the regulation of mammalian embryonic stem cells. Therefore, the miRNA molecules can be used for analyzing regulatory networks of human embryonic stem cells.